

众包系统中基于双向拍卖的诚实多任务分配机制*

崔景妹¹, 孙玉娥^{2,3}, 黄河^{1,3}, 辛煜⁵, 郭寒松^{3,4}, 杜扬^{3,4}



¹(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

²(苏州大学 轨道交通学院, 江苏 苏州 215137)

³(中国科学技术大学 苏州研究院, 江苏 苏州 215123)

⁴(中国科学技术大学 计算机科学与技术学院, 江苏 苏州 215123)

⁵(北京遥感信息研究所, 北京 100011)

通讯作者: 孙玉娥, E-mail: sunye12@suda.edu.cn

摘要: 现有的众包任务分配模型大都假设仅存在一个任务请求者, 难以满足存在多个任务请求者的应用场景. 因此, 该文假设众包系统中存在多个任务请求者, 且每个任务请求者或工人在每轮任务分配中可以请求或完成多个任务, 并在此基础上设计了适用于众包系统的双向拍卖机制, 以激励众包任务被高效地完成. 首先, 将对多个任务感兴趣的工人抽象为多个虚拟工人, 以低报价优先匹配原则, 设计了一个高效任务分配机制; 根据统一定价原则, 为其设计了诚实的定价机制. 然后, 通过理论分析, 证明了所设计的机制可以满足诚实性、个体理性以及收支平衡这3项经济学特性. 最后, 通过仿真实验结果验证了所设计机制的有效性.

关键词: 双向拍卖; 众包; 任务分配; 机制设计; 多物品

中文引用格式: 崔景妹, 孙玉娥, 黄河, 辛煜, 郭寒松, 杜扬. 众包系统中基于双向拍卖的诚实多任务分配机制. 软件学报, 2018, 29(Suppl. (1)): 105-114. <http://www.jos.org.cn/1000-9825/18011.htm>

英文引用格式: Cui JM, Sun YE, Huang H, Xin Y, Guo HS, Du Y. Double auction based truthful multi-unit task allocation mechanism for crowdsourcing systems. Ruan Jian Xue Bao/Journal of Software, 2018, 29(Suppl. (1)): 105-114 (in Chinese). <http://www.jos.org.cn/1000-9825/18011.htm>

Double Auction Based Truthful Multi-unit Task Allocation Mechanism for Crowdsourcing Systems

CUI Jing-Mei¹, SUN Yu-E^{2,3}, HUANG He^{1,3}, XIN Yu⁵, GUO Han-Song^{3,4}, DU Yang^{3,4}

¹(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

²(School of Rail Transportation, Soochow University, Suzhou 215137, China)

³(Suzhou Institute for Advanced Study, University of Science and Technology of China, Suzhou 215123, China)

⁴(Department of Computer Science and Technology, University of Science and Technology of China, Suzhou 215123, China)

⁵(Beijing Institute of Remote Sensing Information, Beijing 100011, China)

Abstract: Most of the existing state-of-the-art studies assume that there is only one task requester for the crowdsourcing systems, which is hard to satisfy the scenario of multiple task requesters. Thus, this work focuses on the crowdsourcing systems consisting of multiple task requesters, where all the requesters and workers can bid or complete multiple tasks in each round. On the basis of that, this paper proposes a truthful double auction mechanism for crowdsourcing systems to incentivize workers and allow efficient task completion. First of all, this method abstracts the workers that interested in multiple tasks into multiple virtual workers, and further designs an efficient task assignment mechanism based on the principle of low bids first matching. A truthful pricing mechanism is developed based on the principle

* 基金项目: 国家自然科学基金(61672369, 61572342, 61873177); 江苏省自然科学基金(BK20161258)

Foundation item: National Natural Science Foundation of China (61672369, 61572342, 61873177); Natural Science Foundation of Jiangsu Province (BK20161258)

收稿时间: 2018-05-01; 采用时间: 2018-08-30

of uniform pricing. In addition, a proof is given to show that the proposed mechanism satisfies three essential economic characteristics: truthfulness, individual rationality, and budget balance through substantial theoretical analysis. Finally, extensive simulations are conducted to study the performance of the proposed auction mechanism, and the simulation results corroborate the theoretical analysis.

Key words: double auction; crowdsourcing; task assignment; mechanism designing; multi-unit task

近年来,随着互联网的飞速发展,众包(crowdsourcing)/群智感知(crowdsensing)作为一种新型计算模式脱颖而出.众包,即通过互联网或移动互联网,将广泛分布在不同地理位置,存在时间与空间隔阂的大众群体互联,利用群体智慧,共同解决复杂问题的一种计算模式^[1].在众包系统中,任务请求者首先将复杂的大问题分解为多个子任务,再通过众包平台(如 Amazon Mechanical Turk^[2])将子任务(sub-task)发布,征集付费网民完成任务.采用众包模式,任务请求者(Task requester/Crowdsourcer)可以充分利用群体智慧,挖掘新的创意方案,解决企业技术难题,降低企业运营成本,提高企业的创新能力^[3].

目前,众包模式已经成功地应用在各个领域,如室内定位^[4]、环境监测^[5,6]、图片标注^[7,8]、数据采集^[9]等.亚马逊旗下的 Mechanical Turk 平台为心理学家提供了超过 50 万的样本供选择,心理学家只用支付给每个样本很少的报酬,即可通过众包的方式开展研究.NASA 众包太空研究项目^[10]——地球上的对照组实验采取众包的形式,让孩子们种植与送往月球一样的植物,减少了数千次实验数据的成本.Foldit 以电子游戏的方式,让用户试着操纵简单的类蛋白质构造,得出实际的蛋白质模型,系统根据折叠的完善程度给予评分.Indiegogo 是美国第二大众筹平台,创意项目可在该平台上筹措资金.除了资金外,这些项目的发起者还能获得重要的信息,这些信息可以帮助项目发起者更深入了解这个产品的市场行情、收集目标用户反馈.然而,众包在广泛应用的过程中,也存在着一定的问题.由于工人完成任务的酬金一般较少甚至最终无酬金可得,导致他们参与众包的积极性严重下降;同时,参与众包项目的工人是不确定的网络人群,他们的技术水平各不相同,提交的数据质量参差不齐,甚至有的工人会提交虚假数据,从而产生冗余信息,导致众包平台不能为任务请求者的数据质量提供保证^[11-13].因此,为了保证众包系统的发展,必须设计以激励(incentives)为基础的任务分配机制,激励更多有能力的网络人群参与,并选择最合适的人群完成任务,从而保证众包项目高效完成.

目前,学术界有关众包系统任务分配机制研究主要针对以下 3 种模型:单请求者-单报价、单请求者-多报价以及多请求者-多报价^[14].对于前两种模型常通过构造单向拍卖(single-sided auction)实现对工人的激励,而第 3 种模型常通过构造双向拍卖(double auction)实现任务交易.在基于单向拍卖的任务分配机制研究中,以单向拍卖构造买卖双方交易过程.如文献[15]针对众包系统的任务分配问题,提出了以平台为中心和以用户为中心的两种众包任务分配模型,两者均以最大化平台效益为目的,前者根据用户的感知时间决定用户的报酬;后者构造拍卖模型,采用贪心技术进行分配并利用净效益进行定价.文献[17]针对某些众包系统不存在常驻用户的问题,考虑用户的流动性,并提出了离线模式(off-line)和在线模式(online)下的两种任务分配模型,前者采用带权二分图匹配作为分配算法并以 VCG 拍卖的方式为工人(worker)计算报酬,后者以贪心算法进行分配并根据寻找关键用户(critical user)进行定价.文献[18]针对移动群智感知中依赖位置感知的问题,提出了一种将 NP-Hard 问题转化为多项式时间可解的 MWSC 问题的一个特例,并设计了相应的近似算法.文献[22]针对众包系统中存在的异质任务分配问题,提出一种将 NP-Hard 问题转化为多项式时间可解的近似分配算法.在基于双向拍卖的众包任务分配机制研究中,文献[16]首次为众包系统考虑了多个任务请求者的情形,并提出了一个基于双向拍卖的任务分配机制,但没有考虑工人的自主选择性,即没有考虑工人因技术水平差异而自主选择能力范围内的任务来完成;文献[19]提出了一种在线的双向拍卖模型,适用于动态模式下的众包系统,但没有考虑工人的多个任务需求;文献[20]也考虑了动态模式下的任务分配模型,但限定工人只满足单任务需求;文献[21]考虑了工人的长期可靠性对平台的影响,但它假定平台只发布同质任务.由于基于双向拍卖的众包任务分配机制研究,均没有考虑到工人的多任务拆分需求,然而这点对于实际中的众包具有重要意义.因此,本文从工人的多任务拆分需求出发,设计一个适用于多任务请求者众包系统的任务分配机制,保证众包项目高效完成.

1 问题建模

本节将给出众包系统中基于双向拍卖的多任务分配模型,并针对所研究的问题模型给出形式化描述.

1.1 系统模型

本文假定所研究的众包系统由多个任务请求者、一个云端的众包平台以及若干工人组成.在每个任务分配周期,请求者依照平台规定,首先向平台发布自己的任务请求方案,包括对请求任务的技术要求、出价及规定数(i.e.,完成任务所需要的最多工人数量)的描述.假设有 m 个请求者在任务分配初期向平台发布请求,用集合 $R=\{1,2,3,\dots,m\}$ 表示这些请求者,用 $S_i=\{t_i,r_i,f_i\}$ 表示请求者 i 的请求方案,其中 t_i 表示任务描述, r_i 表示单任务 t_i 的最高出价,即请求者 i 在一个工人完成一次任务 t_i 后所愿意支付的最大报酬, f_i 表示任务 t_i 规定数,即请求者 i 需要工人完成任务 t_i 所需要的最多工人数量.由于请求者所发布的任务描述各不相同,因而这些任务请求方案中所描述的任务是异质(heterogeneous)的且不可再分割(indivisible)的.在收集完请求者的任务请求方案后,平台将这些请求方案中的任务描述发布在平台上;在工人阅读完这些任务描述后,他们会根据自身能力(ability)情况,选择一些可以做的任务,即一个感兴趣的任務集合,进而制定自己的任务计划并发送至平台.假设系统中有 n 个异质的工人,即他们的兴趣能力等因人而异,我们用集合 $W=\{w_1,w_2,\dots,w_n\}$ 表示这些工人,用 $\theta_j=\{T_j,B_j\}$ 表示工人 w_j 的任务计划,其中 T_j 表示 w_j 申请的任务集,即 w_j 对多个请求者的任务描述符合要求并有兴趣完成, B_j 表示 w_j 对感兴趣任务的任務报价集合.用 $T_j=\{t_i\}_{i \in R}$, $B_j=\{b_{j,i}\}$ 分别表示 w_j 的感兴趣任务集以及对任务集中的任务的报价集.然而,为了最大化自身的利益,某些工人希望获得更多的报酬,因而可能会抬高对任务的报价,进而使得实际开销可能小于对任务的报价.我们用 C_j 来表示 w_j 对 T_j 中每个任务的真实估价(真实开销),其中, $C_j=\{c_{j,i}\}$.

在收集完工人的任务计划后,平台会根据制定的分配和定价算法,选择最合适的请求者和工人,完成工人与任务的匹配,并分别向选择的请求者收取一定的费用,给予完成任务的工人一定的报酬.至此,一个任务分配周期结束.

1.2 问题描述

在众包系统中,任务请求者希望通过支付酬金获取任务资源,工人希望通过完成任务获取一定的报酬,而平台作为中介,在保证自身利益的前提下实现请求者与工人的最优匹配.为了实现这个交易过程,本文构造请求者为买家,工人为卖家,平台为拍卖商的一个双向拍卖模型,进而设计一个保证买卖双方诚实性的双向拍卖机制,满足工人的多任务需求.本文设计的双向拍卖机制,主要包括 3 个部分:构建任务分组,选择请求者和选择工人以及收费和报酬定价计算.

本文用 $x_{j,i} \in \{0,1\}$ 表示平台是否为 w_j 分配任务 $t_i \in T_j$.若 $x_{j,i}$ 取值为 1,表示平台把任务 t_i 分配给 w_j ,即 w_j 成功申请任务 t_i ;若 $x_{j,i}$ 取值为 0,表示平台不将任务 t_i 分配给 w_j ,即工人 w_j 申请任务 t_i 失败.如果工人申请到了某些任务,那么他在完成任务并上传数据后会获得一定的报酬.工人 w_j 获得效益可表示为

$$u_j = \sum_{t_i \in T_j} x_{j,i} p_{j,i} - \sum_{t_i \in T_j} x_{j,i} c_{j,i}, \quad (1)$$

其中, $p_{j,i}$ 表示 w_j 完成任务 t_i 而获得的报酬;若 w_j 没有任何任务申请成功,那么他获得的效益为 0.

在将任务请求者的任务分出去之后,平台根据分配算法选择任务请求成功的请求者 $R_s \subseteq R$.由于请求者之间并无联系,那么每个请求者会以最大化自身利益为前提,从而决定一个任务需要做的最大次数,即任务规定数,这与任务请求者自身建立的效益函数有关.对于请求成功的请求者,收到平台发送的任务数据后,需要向平台支付一定的金额,具体计算由平台采用的定价算法决定.那么任务请求者的效益 u_i^R ,来自完成任务后获得收益与分配任务付出费用之间的差值.

本文的目标是,为众包平台设计一个诚实的双向拍卖的任务分配机制,满足工人的多任务需求,提高任务完成效率,并达到如下 3 点设计目标.

- 诚实性:我们常说,若一个双向拍卖机制是诚实的,那么参与任务拍卖的买家或卖家不能通过虚假报价而提高自身效益.也就是说,在拍卖机制中,诚实报价是买家或卖家的占优策略.

■ 个体理性:每个赢得拍卖的买家或卖家都有非负的效益.也就是对于赢得拍卖的买卖双方来说,买家需要付出的费用不会超过他的出价,卖家得到的报酬不会低于他的报价.因为参与拍卖的请求者和工人,分别会给出自己对任务的出价和报价.对于请求者的出价,是愿意为分配任务出的最大价值,而对于工人的报价,是工人愿意完成任务的最小报酬.

■ 收支平衡:在一个双向拍卖机制中,拍卖商的效益不低于 0,那么这个双向拍卖机制就是收支平衡的.在本文考虑的模型中,平台的效益记为

$$\varphi = \sum_{i \in R} P_i^R - \sum_{j \in W} P_j^W \quad (2)$$

其中, P_i^R 表示对请求者 i 的收费, P_j^W 表示 w_j 获得的报酬.若平台的效益 φ 不低于 0,则保证了收支平衡特性.

2 满足工人多任务需求的双向拍卖机制

首先给出双向拍卖机制的设计细节,主要包括 3 个部分:构建任务分组 G ,选择请求者集 $R, \subseteq R$ 、工人集 $W, \subseteq W$,以及收费与报酬的定价计算;然后,对所设计的拍卖机制进行理论分析,验证本文的拍卖机制:在满足工人多任务需求的前提下,同时也满足 3 条经济特性.

2.1 构建任务分组

本文满足工人的多任务需求且申请每一任务成功与否互不影响.若考虑一个工人只能申请一个任务的情况,那么对于申请多个感兴趣任务的工人,则可以看作是自身的多个虚拟工人.按照这样的思路,本文把每个工人根据自身感兴趣任务集抽象为多个虚拟工人,并且以工人其他的每个感兴趣任务而区分(如,工人 w_1 的任务计划 $T_1 = \{t_1, t_3, t_5\}$,则用 3 个虚拟工人 $\{w_1^1, w_1^3, w_1^5\}$ 表示工人 w_1).这样,虚拟工人 w_j^i 表示工人 w_j 申请了任务 t_i ,以下用 W' 表示虚拟工人集.接下来,我们开始构建任务分组 G ,如图 1 所示.首先将 R 中的请求者按出价非升次序排序,使得 $r_1 \geq r_2 \geq \dots \geq r_m$;然后把排序后的每个请求者 i 对应到任务 t_i ,把所有申请了 t_i 的虚拟工人 w_j^i 加入分组 G_i 和 W' 中;最后得到了 m 个任务分组和 W' ,对不空的 G_i 中所有虚拟工人按报价非降次序排序,使得 $b_{1,j} \leq b_{2,j} \leq \dots \leq b_{q,j}$,至此完成对 G 的构建,具体算法描述见算法 1.

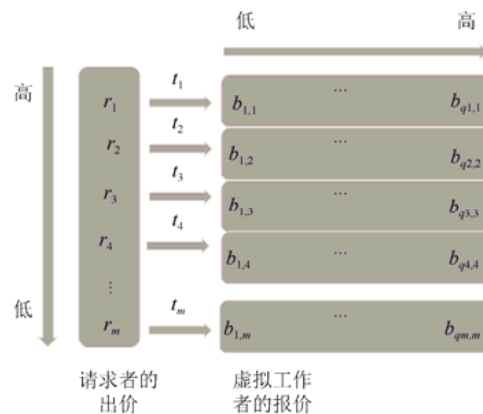


Fig.1 The construction of task groups

图 1 任务分组的构建模型

2.2 分配和定价

G 构建完成后,得到了 m 个分组,请求者 i 对应分组 G_i ,并且在 G_i 中虚拟工人按 $b_{j,i}$ 非降次序排序.接下来,本文将通过分配算法为平台选择可以赢得拍卖的请求者和工人.本文采用高出价低报价优先匹配的原则,为请求者和工人进行匹配,主要包括两个步骤.第 1 步,若在请求者 i 对应的任务分组 G_i 中,满足不大于出价 r_i 和任务规定数 f_i 条件的虚拟工人个数不超过 1,那么请求者 i 请求的任务没有工人申请完成,不能进行任务匹配;反之,

将 i 加入 R_s 中.最后从 R_s 中删除出价最低的请求者 s ,并将 s 的出价 r_s 作为选择虚拟工人的一个约束条件.

第 2 步,在 R_s 中的请求者 i 对应的分组 G_i 中,按照 i 的任务规定数 f_i 以及出价限制 r_s ,若 G_i 中所在序列位置不超过 f_i 且报价不超过 r_s 的虚拟工人个数超过 1,则将最后一个满足条件 w_q^i 之前的虚拟工人加入 W'_s ;否则,从 R_s 中删除请求者 i .

在上面的两个步骤中,通过第 1 步我们确定了请求者 s 和 R_s ,其中我们把 s 称作为关键任务请求者(critical requester);通过第 2 步,我们更新了 R_s ,并得到 W'_s .于是,我们的任务分配就结束了.接下来,对 R_s 、 W'_s 中的用户分别给出收费和报酬的定价计算.

本文采用统一定价技术,对 R_s 中的请求者与 W'_s 中的虚拟工人进行定价计算.由于本文在构建 G 时,把每个 w_j 按任务集 T_j 抽象为多个虚拟工人 w_j^i ,并以 $t_i \in T_j$ 区分抽象得到的多个虚拟工人,那么 w_j 的报酬,就是分配结束后他抽象出的所有 w_j^i 所得报酬 $p_{j,i}$ 的总和.对于分配算法中,本文把每个可以赢得拍卖的 w_j^i 的报酬计为分组中最后一个可以赢得拍卖的 w_q^i 的报价 $b_{q,i}$,再令 w_q^i 输掉拍卖,也就是, w_q^i 对分组所表示的任务 t_i 申请失败.对于没有赢得拍卖的虚拟工人,报酬计为 0.本文对每个赢得拍卖的请求者的收费 r_s ,计为在首轮分配中找到的关键请求者的出价 r_s ,也就是把最后一个赢得拍卖而令其输掉拍卖的请求者的出价作为收费标准.对于没有赢得拍卖的请求者,平台不对他进行收费.于是,得到定价公式(3).

到这里,我们完成了对请求者与工人的分配和定价,具体算法步骤参见算法 2.

$$P_j^W = \sum_{t_i \in T_j} x_{j,i} b_{q,i}, x_{j,i} \in \{0,1\}, p_i^R = r_s \quad (3)$$

算法 1. 构建任务分组 G .

输入:请求者集合 R ,工人集合 W ;

输出:任务分组 G .

- ① $W' = \emptyset, G = \emptyset$;
- ② 对所有请求者按出价非升次序排序,使得 $r_1 \geq r_2 \geq \dots \geq r_m$;
- ③ for each $w_j \in W$
- ④ for each $t_i \in T_j$
- ⑤ $W' = W' \cup w_j^i$;
- ⑥ $G_i = G_i \cup w_j^i$;
- ⑦ end for
- ⑧ end for
- ⑨ for each $G_i \in G, i \in R$
- ⑩ 对 G_i 中 w_j^i 以报价非降次序排序,使得 $b_{1,i} \leq b_{2,i} \leq \dots \leq b_{q,i}$;
- ⑪ end for
- ⑫ return G, W' ;

算法 2. 选择赢得拍卖的用户,并为其定价.

输入:请求者集合 R ,工人集合 W, W' ,任务分组 G ;

输出:赢得拍卖的请求者集 R_s ,赢得拍卖的虚拟工人集 W'_s ,收费集 $\{p_i\}_{i \in R}$ 以及报酬集 $\{p_{j,i}\}_{w_j \in W, t_i \in T_j}$.

- ① $R_s = \emptyset, W'_s = \emptyset$;
- ② for each $G_i \in G$
- ③ if 报价不超出 r_i 的元素个数大于 1
- ④ $R_s = R_s \cup \{i\}, s = i$;
- ⑤ end if
- ⑥ $R_s = R_s \setminus \{s\}$;

```

⑦ end for
⑧ for each  $G_i \in G, i \in R_s$ 
⑨   set  $k=0$ ;
⑩   while  $k < G_i$  的长度 &&  $k < f_i$  &&  $b_k^i \leq r_s$ 
⑪      $k++$ ;
⑫    $k--$ ; /* $k$  表示满足条件的边界用户*/
⑬   if  $k \geq 1$ 
⑭     set  $x_{k,i}=0$ ;
⑮     对前  $k-1$  个用户设置  $x_{j,i}=1$ , 并将其加入  $W'_s$ , 设置  $p_{j,i} = b_k^i, p_i = r_s$ ;
⑯   else  $R_s = R_s \setminus \{i\}$ ;
⑰   end if
⑱ end for
⑲ return  $R_s, W'_s, \{p_i\}_{i \in R_s}, \{p_{j,i}\}_{w_j \in W, i_j \in T_j}$ ;

```

2.3 理论分析

本文提出的双向拍卖机制,满足工人多任务需求性的同时,保证机制满足 3 条经济特性,更符合众包实际应用场景.下面,将给出本文机制的诚实性、个体理性及收支平衡的证明.

本文所设计机制的诚实性主要体现在两个方面:一方面,对于请求者是诚实的;另一方面,对于参与拍卖的工人是诚实的.首先证明本文机制对于请求者的诚实性,我们引入两条定理.

定理 1. 给定包括出价信息的请求者集合以及包括报价信息的工人集合,若请求者 i 以出价 r_i 赢得了拍卖,那么当 i 出价 $r_i' > r_i$ 时,他也一定会赢得拍卖.

由于本文将请求者按出价非升序排列,因此,若 i 提高自己的出价为 r_i' ,不会降低自己在序列中的位置.同时, i 对应的任务分组 G_i 内虚拟工人的报价不变,这样满足出价条件的虚拟工人不会减少,于是当请求者提高了报价后,也一定会赢得拍卖.

定理 2. 给定包括出价信息的请求者集合和包括报价信息的工人集,若请求者 i 以出价 r_i 和 r_i' 赢得了拍卖,那么对请求者 i 的收费相同,或者说, i 不会通过降低自己的出价而减少收费.

由于本文对请求者 i 的定价依赖于首轮分配中的 critical requester s .若 i 以出价 r_i 和 r_i' 均赢得了拍卖,那么平台对 i 的收费相同,均为 r_s .

其次我们需要证明本文提出的机制对于工人是诚实的.也就是说,诚实报价是工人的占优策略,即不诚实报价带来的效益没有诚实报价带来的效益高.而对本文的模型,只需证明对虚拟工人是诚实的.对于参与拍卖的虚拟工人 w_j^i ,我们可以得到如下 4 种关于能否赢得拍卖的报价情况分析.

1) w_j^i 无论诚实或不诚实报价, w_j^i 都会赢得拍卖.由于本文把 w_j^i 所在任务分组 G_i 内最后一个满足条件 (critical requester s 的出价条件及 i 任务规定数条件)且报价最高的虚拟工人的报价 $b_{q,i}$ 作为 w_j^i 的报酬 $p_{j,i}$,所以,即使 w_j^i 诚实或不诚实报价都在赢得拍卖的情况下, w_j^i 的报酬均为 w_q^i 的报价.对于诚实报价而赢得拍卖即真实估价小于报酬,得到的效益大于 0;对于不诚实报价而赢得拍卖,若真实估价大于报酬,效益为负,否则效益仍为报酬减去真实估价.

2) w_j^i 诚实报价赢得拍卖,不诚实报价输掉拍卖.根据我们的分配机制,诚实报价赢得拍卖,说明诚实报价是小于 r_s 的,而得到的报酬一定会大于诚实报价,所以诚实报价可以得到大于 0 的效益;而不诚实报价输掉拍卖,效益为 0.所以这种情况下,诚实报价效益最高.

3) w_j^i 诚实报价输掉拍卖,不诚实报价赢得拍卖.根据我们的分配机制,若诚实报价而输掉拍卖,那么 w_j^i 的诚实报价大于 r_s ,效益为 0;当 w_j^i 降低对任务的估价进行不诚实报价,赢得了拍卖,于是得到序列靠后的 w_q^i 的报

价作为报酬.这种情况下, w_j^i 诚实报价可能刚好作为 w_q^i 而输掉拍卖,那么 w_j^i 的真实估价不小于 w_q^i 的报价即赢得拍卖的报酬.因而不诚实报价,即使赢得拍卖,也会导致 w_j^i 的效益非正.

4) 无论诚实或不诚实报价,均输掉拍卖.在这两种情况下,无论虚拟工人怎样报价都会输掉拍卖,那么得到的效益都为 0.

结合上面 4 点分析,我们可以得到,诚实报价是虚拟工人的占优策略.接下来,我们看一下本文如何体现个体理性以及收支平衡.

根据前文可知,若对请求者的收费不高于其出价而对虚拟工人的报酬不少于其报价,那么本文的机制满足个体理性.由于本文对赢得拍卖的任务请求者的收费为 r_s , 不会超过他们的出价,而对赢得拍卖的虚拟工人的报酬为 $b_{q,i}$ 不会低于其报价,因而,本文机制满足个体理性.最后,由于本文对请求者单任务的收费为 r_s ,而对工人单任务的报酬不高于 r_s ,所以平台得到的效益是不小于 0 的.因而,本文提出的机制保证了收支平衡.

3 仿真模拟与分析

本文提出了一个诚实的双向拍卖机制,并满足工人的多任务需求.在实验仿真部分,实验首先对平台单任务的收益进行分析;其次将从工人满意度与任务完成率(或者成功匹配率)两个角度评定实验效果.为了体现本文算法的优势,本文通过与文献[16]提出的方法(double auction task assignment,简称 DATA)进行对比实验.在我们的实验设置中,请求者的出价均匀分布在区间[50,249],工人的报价均匀分布在区间[1,100].所有的仿真结果均取 1 000 组独立运行结果的算术平均.

3.1 度量标准

本文将工人满意度、任务完成率、平均任务完成率的定义用公式(4)~公式(6)表示.

$$\text{工人满意度} = \frac{\text{有任务可做的人数}}{\text{总工人数}} \quad (4)$$

$$\text{任务完成率} = \frac{\text{分组内赢得拍卖的虚拟工人个数}}{\text{任务规定数}} \quad (5)$$

$$\text{平均任务完成率} = \frac{\text{请求者的任务完成率总和}}{\text{赢得拍卖的请求者个数}} \quad (6)$$

3.2 平台单任务收益分析

在本文提出的双向拍卖机制中,平台作为拍卖商,实现任务交易过程.在这个过程中,平台向成功的请求者收取费用,按单价乘以被做任务数计算;并且平台付给工人报酬.由于每个工人被抽象为多个虚拟工人,因而每个工人的报酬,就是其对应赢得拍卖的虚拟工人获得的报酬总和.如图 2 所示,观察当工人数从 100 变化到 1 000,间隔为 100 时,平台对单任务的平均收费与平均报酬之间的变化关系,进而分析平台的单任务收益.

从图 2 中可以看到,当平台的请求者人数一定时,随着工人数量的增加,平台对请求者单任务平均收费基本处于稳定状态,而对工人单任务的平均报酬数值逐渐下降.这是因为,当请求者数量一定,工人数量增加时,供不应求,工人之间竞争强烈,因而工人的报酬逐渐下降.因此,当工人数量逐渐增大时,平台的单任务收益会越来越高.

3.3 满意度对比分析

如图 3 所示,当请求者数量从 100 变化至 1 000 时,两个算法得到的工人满意度最初随着请求者数量的增加而逐渐上升,而最终趋于平缓;并且本文提出的算法得到的工人满意度较早趋于平缓并高于 DATA 算法的.这是因为随着请求者数目的增加,每个工人被分配到任务的可能性增大,因而,在两个算法中,工人满意度均呈现上升趋势;当请求者增加到可以满足所有工人需求时,工人满意度不再上升.又因为本文考虑工人的多任务需求,而 DATA 仅考虑工人的单任务需求,因而本文工人的满意度高于 DATA.

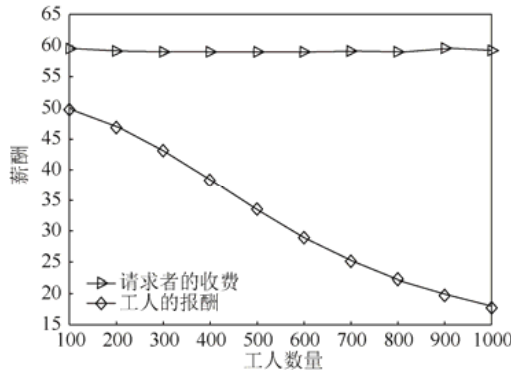


Fig.2 The performance of payment under different number of workers

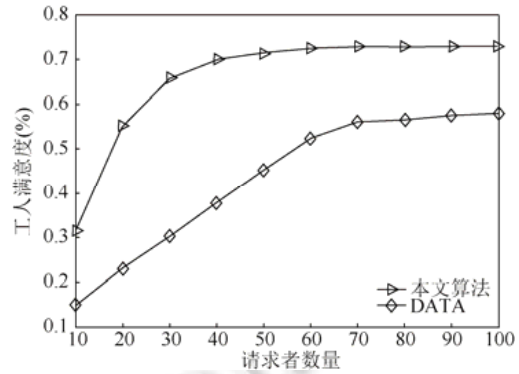


Fig.3 The performance of worker satisfaction ratio under different number of requesters

图2 不同工人数量对请求者收费与工人报酬的影响 图3 不同请求者数量对两个算法的工人满意度影响

在图4中,观察工人数从100变化至1000时两个算法得到的工人满意度的变化趋势.从图4中,我们可以看到,随着工人数量的增加,两个算法的工人的满意度均呈下降趋势且最终趋于平缓;并且本文算法产生的工人满意度一直高于DATA的.这是因为,当工人数量增加,任务供不应求,工人之间竞争激烈,因而其满意度呈现下降趋势;又因为本文算法考虑到工人多任务需求,尽可能使更多的工人满足任务需求,而DATA只考虑了工人的单任务,因而本文算法带来的工人满意度高于DATA算法的.

通过满意度对比分析,可以发现,我们的算法更适用于工人数较多的平台,从而避免分配不足的现象.

3.4 任务完成率分析

在图5中,我们通过比较在请求者的任务规定数限制在50以内和100以内的两种情况下,随着工人数量的增加,请求者的平均任务完成率的变化规律,来分析本文机制的实用性能.从图5中,我们可以看到,随着工人的数量增加,两种情况下的平均任务完成率均呈现上升趋势,且在初期,规定数为50情况下的任务完成率上升更快并高于规定数100情况下的;当工人数量饱和时,两种情况下的任务完成率均不再增加而趋于平缓,并且前者比后者更早趋于平缓.这是因为,请求者的任务完成率与满足条件的工人个数、任务规定数有关,当工人数量较少时,任务规定数越大的请求者对满足条件的工人的需求越大,因而任务完成率低于任务规定数较小的请求者,并且随着工人的数量增加均呈现上升趋势;当工人数量充足时,请求者的任务需求都尽可能被满足,并且规定数小的请求者更先于被满足,因而,两种情况下的任务完成率最终趋于平缓并且任务完成率高达93%.

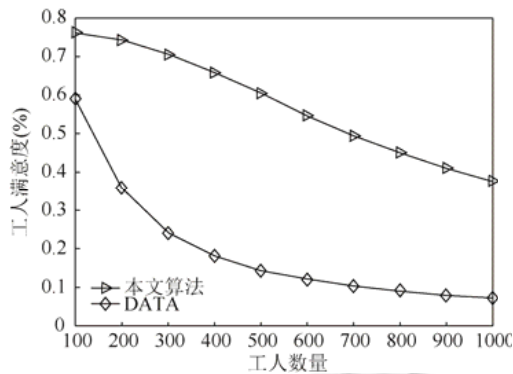


Fig.4 The performance of worker satisfaction ratio under different number of workers

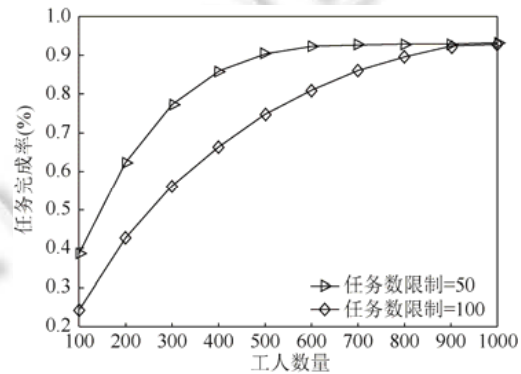


Fig.5 the performance of task completion ratio under different number of workers and task limit

图4 不同工人数量对两个算法中工人满意度的影响

图5 不同工人数量以及不同最大任务规定数限制对任务完成率的影响

4 总 结

基于双向拍卖的多任务分配模型,本文提出了一种满足工人多任务需求的双向拍卖机制.该机制在请求者限定任务数的基础上,通过双向拍卖,实现请求者与工人之间的任务交易过程,保证了参与者的诚实性,在激励用户的同时满足用户的高参与度并且保证了高达 93%的任务完成率,具有实用价值.

References:

- [1] Howe J. The rise of crowdsourcing. *Wired*, 2006,14(14):1–5.
- [2] <https://www.mturk.com/>
- [3] Brabham DC. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence the Int'l Journal of Research Into New Media Technologies*, 2008,14(1):75–90.
- [4] Rai A, Chintalapudi KK, Padmanabhan VN, *et al.* Zee: Zero-Effort crowdsourcing for indoor localization. In: *Proc. of the Mobicom 2012*. 2012. 293–304.
- [5] Ballesteros J, Rahman M, Carburnar B, *et al.* Safe cities. A participatory sensing approach. *Local Computer Networks, IEEE*, 2012. 626–634.
- [6] Rana RK, Chou CT, Kanhere SS, *et al.* Ear-phone: An end-to-end participatory urban noise mapping system. In: *Proc. of the ACM/IEEE Int'l Conf. on Information Processing in Sensor Networks (IPSN)*. 2010. 105–116.
- [7] Deng J, Dong W, Socher R, *et al.* Imagenet: A large-scale hierarchical image database. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2009. 248–255.
- [8] Torralba A, Fergus R, Freeman WT. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008,30(11):1958–1970.
- [9] Demartini G, Difallah DE, Cudré-Mauroux P. ZenCrowd: Leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In: *Proc. of the Int'l Conf. on World Wide Web. ACM*, 2012. 469–478.
- [10] <https://www.nasa.gov/>
- [11] Munro R. Crowdsourcing and the crisis-affected community. *Information retrieval*, 2013,16(2):210–266.
- [12] Lebraty JF, Lobre-Lebraty K. *Crowdsourcing: One Step Beyond*. Wiley-ISTE, 2013. 1–119.
- [13] Ye H, Kankanhalli A. Solvers' participation in crowdsourcing platforms: Examining the impacts of trust, and benefit and cost factors. *Journal of Strategic Information Systems*, 2017,26(2).
- [14] Zhang X, Xue G, Yu R, *et al.* Truthful incentive mechanisms for crowdsourcing. In: *Proc. of the Computer Communications (COMCOM). IEEE*, 2015. 2830–2838.
- [15] Yang D, Xue G, Fang X, *et al.* Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing. In: *Proc. of the Int'l Conf. on Mobile Computing and Networking (Mobicom 2012)*. ACM, 2012. 173–184.
- [16] Xu W, Huang H, Sun YE, *et al.* DATA: A double auction based task assignment mechanism in crowdsourcing systems. In: *Proc. of the Int'l ICST Conf. on Communications and Networking*. 2014. 172–177.
- [17] Feng Z, Zhu Y, Zhang Q, *et al.* Towards truthful mechanisms for mobile crowdsourcing with dynamic smartphones. In: *Proc. of the IEEE ICDCS*. 2014. 11–20.
- [18] Feng Z, Zhu Y, Zhang Q, *et al.* TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing. In: *Proc. of the IEEE INFOCOM*. 2014. 1231–1239.
- [19] Wei Y, Zhu Y, Zhu H, *et al.* Truthful online double auctions for dynamic mobile crowdsourcing. In: *Proc. of the IEEE INFOCOM*. 2015. 2074–2082.
- [20] Qin H, Zhang Y, Li B. Truthful mechanism for crowdsourcing task assignment. In: *Proc. of the IEEE Int'l Conf. on Cloud Computing*. 2017. 520–527.
- [21] Wang H, Guo S, Cao J, *et al.* MeLoDy: A long-term dynamic quality-aware incentive mechanism for crowdsourcing. In: *IEEE Trans. on Parallel & Distributed Systems*, 2018,29(4):901–914.
- [22] Cui JM, Sun YE, Huang H, *et al.* TCAM: A truthful combinatorial auction mechanism for crowdsourcing systems. In: *Proc. of the WCNC*. 2018. 1–6.



崔景妹(1993-),女,安徽省蚌埠人,硕士生, CCF 学生会员,主要研究领域为众包.



孙玉娥(1983-),女,博士,副教授,CCF 专业会员,主要研究领域为无线通信网络,博弈论,交通流量统计.



黄河(1983-),男,博士,副教授,CCF 专业会员,主要研究领域为无线通信网络,网络经济学,流量统计.



辛煜(1988-),男,博士,工程师,主要研究领域为遥感大数据信息挖掘.



郭寒松(1990-),男,博士生,主要研究领域为众包,机器学习.



杜扬(1992-),男,博士生,CCF 学生会员,主要研究领域为众包,真值发现,机器学习.

www.jos.org.cn

www.jos.org.cn